Class-Guided Building Extraction from Ikonos Imagery*

D. Scott Lee, Jie Shan, and James S. Bethel

Abstract

Recent high-resolution satellite images provide a valuable new data source for geospatial information acquisition. This paper addresses building extraction from Ikonos images in urban areas. The proposed approach uses the classification results of Ikonos multispectral images to provide approximate location and shape for candidate building objects. Their fine extraction is then carried out in the corresponding panchromatic image through segmentation and squaring. The ECHO classifier is used for supervised classification while the ISODATA algorithm is used for unsupervised classification and subsequent image segmentation. The classification performance is evaluated using the classification confusion matrix, while the final building extraction results are assessed based on the manually delineated results. A building squaring approach based on the Hough transformation is developed that detects and forms the rectilinear building boundaries. A number of sample results are presented to illustrate the approach and demonstrate its efficiency. It is shown that about 64.4 percent of the buildings can be detected, extracted, and accurately formed through this process. Remaining difficulties are high percentage false alarm errors caused by the misclassification of road and building classes as well as occlusion and shadows that may mislead the extraction process.

Introduction

Recent successfully launched satellites (Ikonos, EROS, QuickBird) provide high resolution (around 1 meter or better) images that can be used for the extraction of geospatial features for topographic mapping applications (Petrie, 2001). Both the potential and methodology for this objective have been studied over the last two years. Li *et al.* (2000) and Muller *et al.* (2001) used simulated Ikonos imagery to investigate the obtainable accuracy from precision photogrammetric processing. Dial (2000) reported on the investigation on the potential of Ikonos images and conclude that the accuracy requirement for large scale (1:4,800) topographic maps can be met with Ikonos images. Baltsavias *et al.* (2001a) present a comprehensive evaluation of the radiometric and geometric property and accuracy of the standard Ikonos product. Fraser *et al.* (2001; 2002) used different forms of geometric transformations to further rectify the standard Ikonos image product and attempt to reach a sub-meter accuracy. Toutin (2001) used 13 panchromatic and multispectral Ikonos Geo-product images over seven study sites to study the error propagation in the bundle adjustment and ortho-rectification process. It was found that a 2- to 4-meter accuracy in the ortho-rectified Ikonos image can be achieved if a proper DEM (digital elevation model) is available. A study on the use of stereo Ikonos imagery shows that 5- to 10-m contour lines can be derived with the highest topographic standard (Toutin *et al.*, 2001). As for geospatial feature extraction for topographic mapping, Baltsavias *et al.* (2001a) and Fraser *et al.* (2001; 2002) present results on building extraction from Ikonos stereo images. A comparative study with the results obtained from aerial photographs concludes that about 15 percent of the building areas, as measured in aerial images, cannot be modeled using Ikonos images. An assessment based on 19 GPS surveyed check points at roof corners suggests that the Ikonos-derived building model can reach an accuracy of better than 1 meter both in planimetry and elevation (Baltsavias *et al.*, 2001a). Sohn and Dowman (2001) used a local Fourier transformation to analyze the dominant orientation angle in a building cluster and extract rectilinear building outlines from Ikonos imagery based on a binary-space partitioning tree. Dial *et al.* (2001) present an investigation on automated road extraction in wide suburban roads.

Regarding the fundamental methodology in geospatial feature extraction from aerial and space images, abundant experience has been gained in the past few years. A collection of state-of-the-art articles can be found in the periodical proceedings edited by Grün *et al.* (1995), Grün *et al.* (1997), and Baltsavias *et al.* (2001b). Mayer (1999) presents a comprehensive survey on the techniques used for image-based building extraction. These collections and surveys suggest a common understanding that multiple data sources and external knowledge or models about the buildings are needed to reach an automated solution. Reported methodologies include, only naming a few as representative, the use of additional information, such as topographic maps (Haala, 1999; Förstner and Pluemer, 1997), laser ranging or elevation data collected from lidar or stereo pairs (Baillard and Maitre, 1999; Hofmann, 2001; Kim and Muller, 2001; Weidner and Förstner, 1995), multispectral and hyperspectral images (Haala, 1999; Huertas *et al.*, 1999; Huertas *et al.*, 2000; McKeown *et al.*, 1999), as well as various building models (Braun *et al.*, 1995; Förstner and Pluemer, 1997; Förstner, 1999).

This paper is focused on automated building extraction guided by classification results using both multispectral and panchromatic Ikonos images. The approach developed consists of the following steps. First, a supervised classification is undertaken for the four-band multispectral Ikonos images with focus on building detection. After the classification results are vectorized, the location and extent of the building objects are used to define a working window over the corresponding panchromatic image for precision delineation and extraction. In order to avoid the difficulty in determining the threshold for

*This paper is an enriched and revised version of a paper presented at the 2002 ASPRS Annual Conference.

Geomatics Engineering, School of Civil Engineering, Purdue University, West Lafayette, IN 47907-1284. Corresponding author, Jie Shan (jshan@ecn.purdue.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 69, No. 2, February 2003, pp. 143–150.

0099-1112/03/6902–143$3.00/0

image binarization, an unsupervised classification approach is used to separate building from background in the panchromatic image. Once building boundaries are segmented, squaring building polygons based on the Hough transformation is implemented such that the extracted buildings will be composed of rectilinear outlines. The performances of classification and building extraction are respectively assessed based on manually collected building delineations.

The remainder of the paper is organized as follows. Following the description of the test data, the next section presents buildings detected from the supervised classification. Its performance is evaluated visually and by using the confusion matrix. Then building segmentation using the ISODATA (Iterative Self-Organizing Data Analysis Technique) unsupervised classification approach and building squaring based on the Hough transformation are described. The efficiency of building segmentation and squaring is illustrated by a number of sample results in this section. This is followed by comprehensive statistics to evaluate the performance of the automated approach with reference to the manually delineated buildings. The final section contains concluding remarks and prospects for future work.

Test Data and Classification

The study area is located in the surroundings of Camp Lejeune, North Carolina. Multiple Ikonos panchromatic and multispectral images of the entire region are available. The images used in this study are a single mosaicked Ikonos four-band multispectral image and the corresponding Ikonos panchromatic image. The images were geo-rectified by the image vendor Space Imaging, Inc. Plate 1 shows the study area of 5.8 by 3.6 square kilometers centered on Camp Lejeune. The Ikonos image is displayed in Plate 1a with the IR (infrared) band for the red channel, the green band for the green channel, and the blue band for the blue channel. The properties of the Ikonos images are listed below:

- Spectral resolution: Four bands (near IR/R/G/B), 11 bits/pixel
- Spatial resolution: Four meters/pixel (multispectral) and 1 meter/pixel (panchromatic)
- Region: About 5800 by 3600 square meters
- Preprocessing from Space Imaging, Inc.: Standard Geometrically Corrected (GEO), mosaicked
- Horizontal positional accuracy (root-mean-square error): 25 meters
- Map projection: UTM Zone 18, WGS-84

The objective of the classification was to obtain approximate locations and shapes of potential buildings based on their spectral characteristics in the Ikonos multispectral images. Keeping this in mind, the classification process was designed and carried out with the primary concern in capturing buildings. First, the classification process was focused on the urban portion of the multispectral image where buildings were concentrated, whereas other geographic features such as water, sand, and vegetation were less emphasized. Second, special attention was given to roof and road material classes in terms of defining subclasses and selecting an adequate number of training samples, because our experience has shown that these two classes often have higher confusion rates than other classes in the classification results. Table 1 lists the feature classes, subclasses, and number of training pixels selected for the classification. Third, instead of using the ordinary maximum-likelihood classifier that takes into account only the spectral properties of the geographic features, our classification uses the ECHO (Extraction and Classification of Homogeneous Objects) (Kettig and Landgrebe, 1976) classifier, which incorporates spectral as well as spatial variations into the classification. While a pixel classifier assigns each individual pixel to one of the classes, ECHO first segments the image into groups of statistically similar, contiguous pixels (i.e., region growing). A maximum-likelihood classifier is then used to assign a single class to all pixels belonging to the same group. ECHO can often provide higher accuracy than an ordinary pixel classifier and, because a single class is assigned to a homogeneous region, the results typically contain fewer speckles. For these reasons, it is expected that the ECHO classifier results in vectorized roof polygons that better approximate building boundaries. The Multispec software (Biehl and Landgrebe, 2002) with ECHO classifier was used to perform the classification. It is expected that these steps will maximize the potential of Ikonos multispectral images for building detection and consequently simplify the subsequent building extraction process using the panchromatic image. Plate 1b shows the thematic map resulting from the ECHO classification for the original Ikonos image in Plate 1a. A close-up of the classification results of a highly built-up area is shown in Plate 2a with its roof/building class overlaid on the corresponding panchromatic image in Plate 2b.

TABLE 1. CLASSES AND NUMBER OF TRAINING PIXELS

Project Class Name	Number of Training Pixels
Roof	
- Roof Gray	309
- Roof Light Gray	307
- Roof Dark Gray	643
- Roof White	404
- Roof Blue	841
Road	
- Road Gray	1061
- Road Dark Gray	161
- Road White	250
Water	652
Grass	519
Trees	232
Bare Soil	181
No Data	361
Total	5921

In order to quantitatively assess the classification performance, Table 2 presents the confusion matrix (Mather, 1999, Chapter 8) for the classification results. The class names in the left column of the table stand for the ground truth, while the names along the top are for the classification results obtained with the ECHO classifier. The number of pixels assigned to each class through classification is listed in the corresponding cell. Ideally, the confusion matrix should be a diagonal matrix with all its off-diagonal elements being equal to zero. Any non-zero off-diagonal element indicates the existence of misclassification. As is shown in Table 2, the misclassification occurs primarily between only two subclasses, *Roof Gray* and *Road Gray*, with classification accuracies of 77.3 percent (248/321) and 90.7 percent (976/1076), respectively. These are the two lowest accuracies of the 13 feature classes. For almost all other classes, the misclassification rates are zero. The overall (average) training class accuracy achieved by the ECHO classification is 96.5 percent.

A visual evaluation of Plate 2 is also helpful for assessing the classification results. Comparing Plate 2a with Plate 2b reveals that the classification process correctly separated the classes of trees, grass, buildings, and roads with high accuracy, as indicated in Table 2. As for the primary interest, the majority of building pixels (greater than 90 percent) were correctly identified and they corresponded to the general building shapes on the ground. However, a visual check of Plate 2b also suggests that a large number of non-building pixels were incorrectly labeled as buildings (false positive/alarm errors). Most of these pixels were road pixels and form a lot of small, irregular speckles scattered over the entire image. Some of them appear in the vicinity of buildings and severely change the building shapes. All this makes it impossible for the building class to be

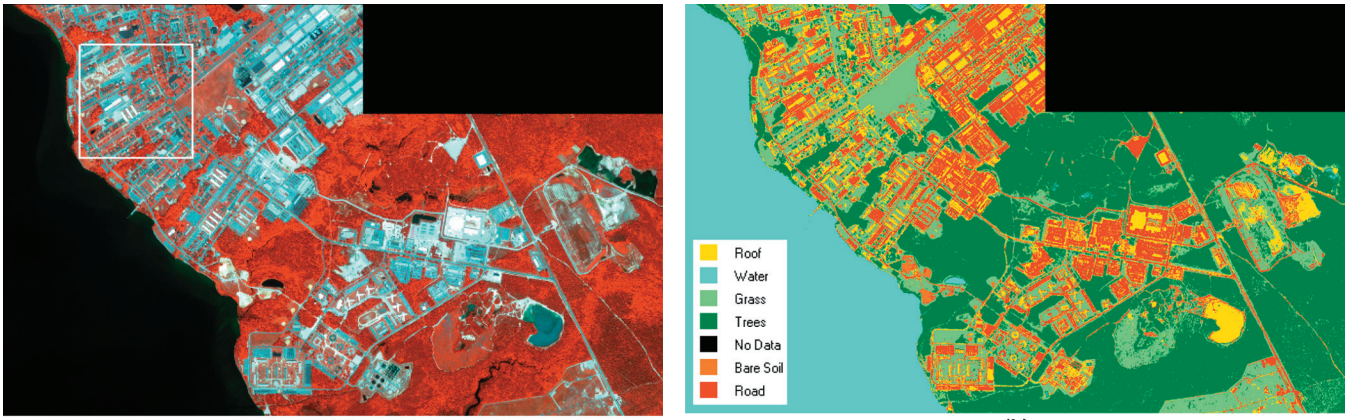


Plate 1. Ikonos multispectral image and its classification (Camp Lejeune, North Carolina). (a) Multispectral image (IR, G, B combination). (b) Class map obtained from the ECHO algorithm.

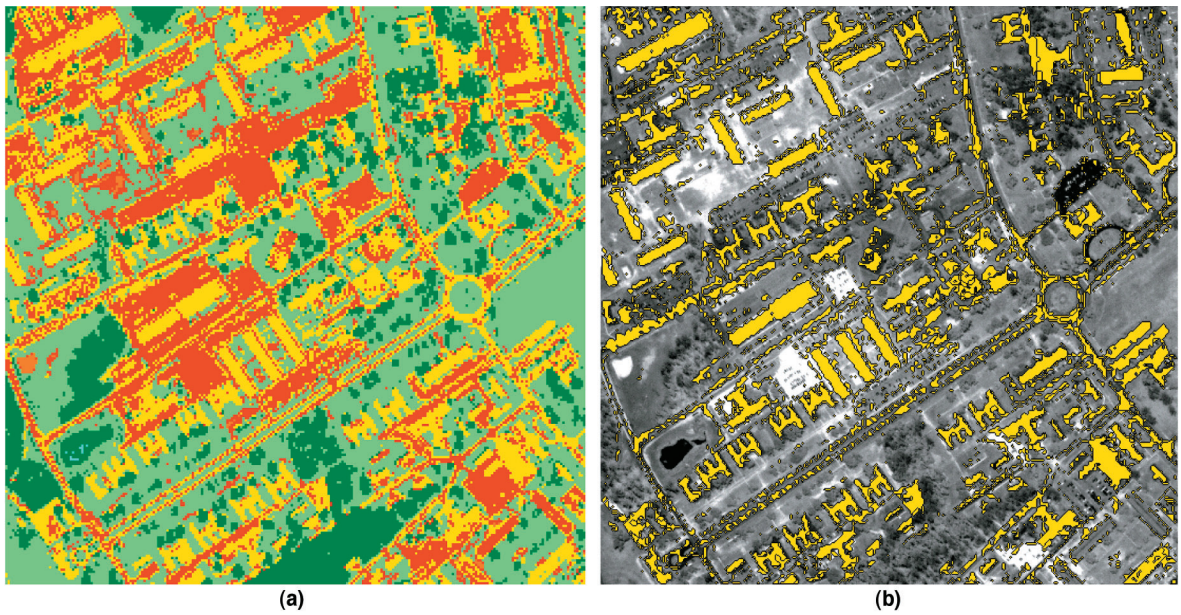


Plate 2. A close-up of the classification results in a highly built up area. (a) Class map (see Plate1b for legend). (b) Building class overlaid on the pan. image.

directly used without further processing. The task in the subsequent steps is essentially to screen out the non-building pixels and refine the correctly identified buildings.

Building Extraction

The objective of building extraction is to obtain accurate building boundaries using the roof class resulting from the above classification process. As shown in Figure 1, this task consists of several steps, which will be described in the following subsections.

Cleaning of Classification Results

The roof class is first converted to vector form so that each potential building will then be represented as a polygon object. The areas are then calculated for each building polygon. Building polygons below a given area threshold are regarded as speckles caused by misclassification and eliminated from further processing. In our study, a threshold for building polygon areas was set as 50 square meters on the ground, or 1.75 by 1.75 pixels on the multispectral image. Most of the small speckle in Plate 2 can then be eliminated from further processing.

Building Segmentation

The task of this step is to determine the building boundary. It is performed in the panchromatic image guided by the building polygons obtained from the classification results. The extent of the building polygons is first used to define the working window for segmentation. Instead of using a conventional binary image segmentation approach, e.g., Otsu's method (Otsu, 1976), we propose to use an unsupervised classification to segment the (panchromatic) image into several (greater than two)

TABLE 2. CONFUSION MATRIX FOR THE CLASSIFICATION RESULTS

Project Class Name	Number of Samples in Thematic Image Class												
	1 Rf Lt Gr	2 Rf Gr	3 Rf Dk Gr	4 Rf Wh	5 Rf Bl	6 Water	7 Grass	8 Trees	9 No Data	10 Bare S	11 Rd Dk Gr	12 Rd Gr	13 Rd Wh
Roof Lt Gray	304	0	0	0	0	0	0	0	0	0	0	0	3
Roof Gray	0	248	2	0	0	0	0	0	0	0	0	59	0
Roof Dk Gr	0	5	597	0	0	0	0	0	0	0	0	41	0
Roof White	0	0	0	402	0	0	0	0	0	1	0	0	1
Roof Blue	0	0	0	0	839	0	0	0	0	0	0	0	2
Water	0	0	0	0	0	652	0	0	0	0	0	0	0
Grass	0	0	0	0	0	0	518	1	0	0	0	0	0
Trees	0	0	0	0	0	0	3	229	0	0	0	0	0
No Data	0	0	0	0	0	0	0	0	361	0	0	0	0
Bare Soil	0	0	0	0	0	0	0	0	0	181	0	0	0
Road Dk Gr	0	0	0	0	0	0	0	0	0	0	161	0	0
Road Gr	0	68	17	0	0	0	0	0	0	0	0	976	0
Road White	3	0	0	0	0	0	0	0	0	0	0	0	247
TOTAL	307	321	616	402	839	652	521	230	361	182	161	1076	253
Accuracy (%)	99	77.3	96.9	100	100	100	99.4	99.6	100	99.5	100	90.7	97.6

Overall Class Performance (5715/5921) = 96.5%

Highlighted areas show pixels misclassified as roof (lower left) and road (upper right)

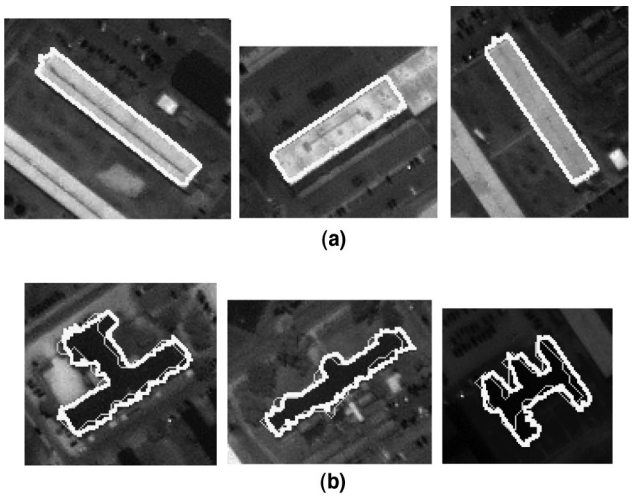


Figure 1. Building polygons obtained from classification (thin) and segmentation (thick) for high contrast (a) and low contrast (b) areas. The classification was done using ECHO and the segmentation was done using ISODATA.

classes. In this way, multiple gray values in the working window can then be taken into account and the difficulty in determining the segmentation threshold is avoided. After the image is segmented into several (usually three to five) classes, the segmented polygon that has the maximum area of intersection with the building polygon obtained from the classification will be selected as the building delineation result from the panchromatic image.

The unsupervised classification approach to segmentation used in this study is the ISODATA algorithm (Mather, 1999, Chapter 8). It allows for the specification of the minimum and maximum number of output classes, maximum number of iterations, class merge distance (in gray values), maximum number of classes that can be merged in a single iteration, minimum number of pixels necessary to form a class, and percentage of pixels which can change classes each iteration. The flexibility given by these different parameters, especially the specification of more than two output classes, helps to separate buildings from multiple non-building backgrounds and results in satisfactory segmentation results. This approach is especially beneficial in low contrast areas where building and background have similar gray values. Figure 1 presents the building delineation results. As is shown, both classification and segmentation can yield equally fairly good results for high contrast buildings (upper figure), whereas the segmentation results (thick line) are more accurate than the classification results (thin line) for low contrast images (bottom figure).

Building Squaring

Building squaring is used to refine and shape the building delineation results from the segmentation. In this study, the building model of rectilinear boundaries is used, namely, the buildings are composed of straight lines that meet at 90-degree angles. This simple building model applies for almost all the buildings in the test area as shown in Plate 1a (overview) and in Figures 1 and 2 and Plate 3a (zoomed-in). The task in this step is therefore to detect the straight building boundaries and enforce the right-angle condition on them, i.e., to square the building corners. As shown in Figure 1a, in situations where good contrast between building and background exists, the building boundaries obtained from both image classification and segmentation can yield fairly good straight lines. In the majority of cases, however, delineations based on image segmentation result in very irregular building boundaries as shown in Figure 1b; even segmentation can not greatly improve the classification results.

Several methods exist that generate straight edges from the irregular boundaries (Ray and Ray, 1992; Costa and Cesar, 2001, Chapter 5). These are primarily polygon filtering methods that seek to eliminate redundant boundary points based on distance and/or area thresholds. The polygon is reduced to its most dominant points depending on threshold settings, resulting in a few straight edges that approximate the original polygon. Although these methods produce desirable refinements, they do not enforce a "square" building model. To enforce the "squaring" criterion, a procedure is developed in this study based on the Hough transformation described, e.g., in Gonzalez and Woods (1993, pp. 432–438).

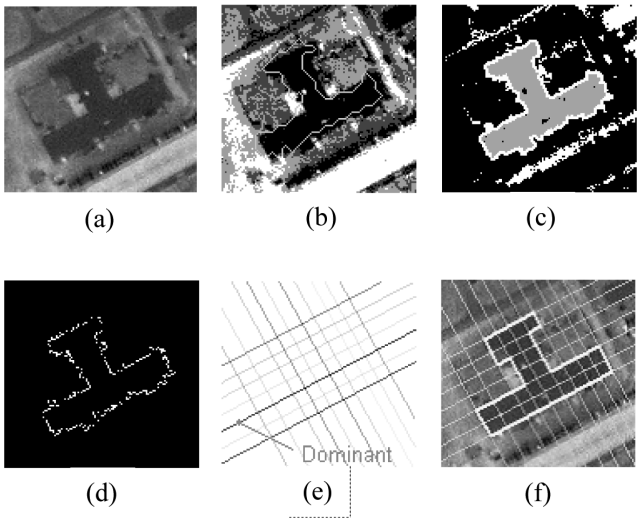


Figure 2. Building squaring using the Hough transformation. (a) Processing window on the panchromatic image. (b) Image segmentation result with the class polygon shown by thin lines. (c) Binarized image. The contour having largest area of intersection with the class polygon is highlighted in thick line. (d) Contour points input to the Hough transformation. (e) Straight lines extracted from contour points. The darker the line, the more points making up the line. (f) Rectangular cells formed from extracted lines were thresholded and traced to create the final building delineation (thick line).

The procedure developed for building squaring uses the contour points derived from image segmentation as input. While these boundaries tend to be irregular, it is reasonable to assume that a number of those boundary points actually lie, within a certain threshold, on the straight edges formed by the building walls. In our approach, the first step is to determine the major orientation of the building by finding the dominant line from the boundary points. The dominant line is defined as the straight line composed of the greatest number of points and is determined by counting the number of points returned by the Hough transformation. Next, all lines parallel and perpendicular to the dominant line are obtained. Because the Hough space is discrete, lines with orientation closest to being orthogonal or parallel are chosen, respectively. At this point, a high density of parallel and perpendicular lines is possible. To eliminate some of the redundant lines, a minimum building side length is defined, forcing a minimum spacing between the lines. The parallel and perpendicular lines are sorted in descending order by the "strength" (number of points on the line). The strongest line (the dominant line) is retained, and then each parallel line in order of descending strength is tested and retained only if it is outside the minimum spacing of a line that has already passed the test. In this way the lines are thinned but with a bias toward retaining the strongest lines. The procedure is repeated for the set of perpendicular lines. After this step, the remaining lines will form a grid that aligns with the walls of the building. The grid forms a set of rectangular cells. In the third step, each cell is overlaid upon the binary image (from segmentation) and the percentage of "white" building pixels within each cell is computed. A cell is retained if it contains a specified percentage (40 percent) or more of white pixels. The final delineation is obtained by tracing the boundary of the retained cells. The tracing step retains only the corner points of the building, i.e., the minimum set of points necessary to define the 2D building contour. The result is a building outline composed entirely of orthogonal edges. Figure 2 illustrates the process of building squaring using the Hough transformation. Sample results are shown in Plate 3a, where both manual (red) and automated (yellow) extraction results are shown. Plate 3b presents all the manually delineated (red) and automatically extracted (yellow) buildings over a 1- by 1-km assessment area. The performance of the automated extraction approach will be evaluated in the next section.

Quality Control and Assessment

To independently and comprehensively assess the overall performance of the entire building extraction procedure, a 1-kilometer square area within the working area was selected for the evaluation (shown by the upper-left box in Plate 1a). Within that area, 112 buildings were manually found and delineated. This manual delineation served as the reference with which to compare the automatic extraction results.

The quality of the automatic process was evaluated for the two consecutive steps—classification and segmentation followed by squaring. To do so, quality measures first needed to be defined. We used manually delineated buildings as a reference or ground truth to assess the automatic process. The automatically extracted buildings (either from classification or squaring) and the reference buildings were compared pixel-by-pixel. The results for each pixel in the image fell into one of four categories (McKeown *et al.*, 1999):

True Positive (*TP*): Both automated and manual methods label a pixel as building.
True Negative (*TN*): Both automated and manual methods label a pixel as background (non-building).
False Positive (*FP*): Only the automated method labels a pixel as building.
False Negative (*FN*): Only the manual method labels a pixel as building.

Using these four categories, the following statistical measures are computed to evaluate the performance of the automated building extraction process (McKeown *et al.*, 1999):

$$\textbf{Branching Factor: } \frac{FP}{TP}$$

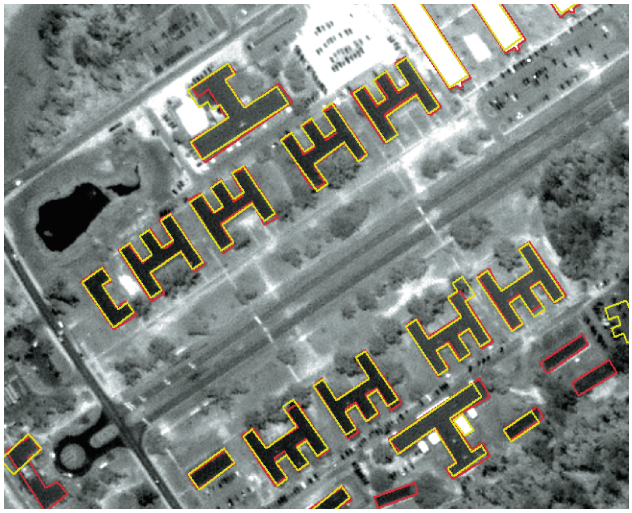
$$\textbf{Miss Factor: } \frac{FN}{TP}$$

$$\textbf{Building Detection Percentage: } 100 \cdot \frac{TP}{(TP + FN)}$$

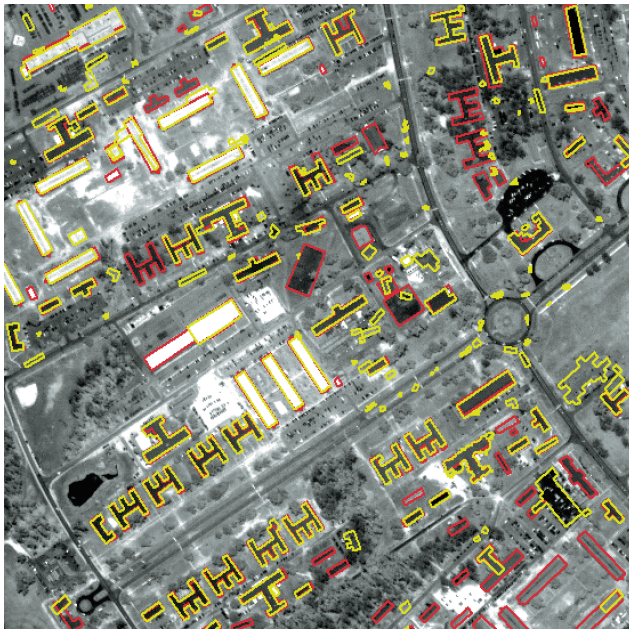
$$\textbf{Quality Percentage: } 100 \cdot \frac{TP}{(TP + FP + FN)}$$

The interpretation of the above calculation is as follows. The branching factor indicates the rate of incorrectly labeled building pixels (*FP*), while the miss factor gives the rate of missed building pixels (*FN*). These two factors are represented with respect to the correctly identified building pixels (*TP*). They describe the two types of potential mistakes (*FP* and *FN*) that may occur in the automatic process. The building detection percentage, where the denominator ($TP + FN$) stands for the total number of true building pixels, indicates the percentage of building pixels correctly labeled by the automatic process. The quality percentage, where the denominator stands for the sum of the total number of true building pixels ($TP + FN$) and the number of wrongly labeled building pixels (*FP*), describes how likely a building pixel produced by the automatic approach is true, and is the most stringent measure of the overall results of the four statistics.

The objective of the quality control is to reduce the two types of possible errors (*FP* and *FN*) in the automatic process.



(a)



(b)

Plate 3. (a) A close-up sample of building squaring results (red: manual delineation; yellow: automatic extraction). (b) Buildings in the 1- by 1-km area used for the performance assessment. Automatically delineated buildings (yellow line) for this area are compared against the 112 manually delineated buildings outlined here with red lines. For a clearer visualization, polygons with less than 100 total pixels are not displayed.

First, the more training sites selected, the fewer errors will occur, however, at the cost of more manual and interactive operations. In our study, the selection of training sites is made realistic. As is shown in Table 1, a total of 5921 pixels were selected as training sites, which is about 0.5 percent of the entire image. This balances the amount of manual training and the expected representative property of the selected training sites. The second quality control is implemented in the segmentation process. Every segmentation polygon (on the panchromatic image) within the working window defined by the classification polygon undergoes a consistency check with the classification polygon (on the multispectral image). When either the areas or the locations of the classification and segmentation polygons are too dissimilar, the segmented results are considered unlikely to be correct and are dropped. Two tests were conducted in this process: (1) the common (intersection) area of the two polygons is larger than 10 percent of the area of the classification polygon, and (2) the common (intersection) area of the two polygons is larger than 40 percent of the area of the segmentation polygon. If either test fails, the segmentation and classification polygons are regarded as inconsistent and the segmentation polygon will be discarded. This test is continued for each segmentation polygon within the working window. The segmentation polygon that passes the above two tests and has the maximum overlap with the classification polygon will be retained as a building polygon for squaring process. The third quality control is to eliminate small speckle polygons produced in the classification. In our study, the area threshold was set up as about 7 by 7 square meters on the ground. Our experience shows that the entire quality control process can significantly reduce the false positive error or branching factor from 1.96 to 0.39 (see Table 3).

The segmentation and squaring procedure, as expected, can refine the initial building shapes obtained from the classification. As shown in Figure 3a, although buildings may only be partially (about 40 percent) captured in the classification step, the segmentation and squaring step can still successfully delineate the entire building. The results are satisfactory. However, misclassification in the supervised classification step can result in insufficient information for building segmentation. As is shown in Figure 3b, only small disconnected speckles for a large building were correctly labeled. This does not provide sufficient information for the subsequent segmentation and squaring step to "pull in" the building image for refinement. As a result of this, the three buildings were missed in the final extraction as shown in Figure 3b.

Table 3 shows the four overall quality measures calculated, respectively, for the classification results and the final segmentation/squaring results. The branching factor and missing factor columns, respectively, give the rates of *FP* and *FN* errors in the classification and segmentation steps. A few observations can be made on this. First, as is shown in this table, both error rates cannot reach minimum at the same time. Reducing the branching factor (from 1.96 to 0.39) is at the cost of boosting the miss factor (from 0.38 to 0.56). This is because true building speckles will also be eliminated when the quality control process removes speckles that are too small either to be buildings or to successfully guide the segmentation and squaring process to extract the entire building. Figure 3b is an example of this scenario where true building speckles are not informative enough to guide the segmentation. Given the fact that small building speckles do not help in the final complete building extraction, they are sacrificed to obtain a much better false positive detection rate (the branching factor from 1.96 to 0.39). Otherwise, a tedious manual interaction will be needed to edit the results where most extracted polygon objects are not buildings, which apparently will make such an automatic process practically useless. Second, the miss factor in classification (0.38) dominates the miss factor of the entire automatic process. As shown in Figure 3, although many buildings that are partly detected in the classification process can afterwards be completely extracted in the segmentation and squaring process, many buildings will still be missed due to the limitation of image quality and the limited multispectral bands of the Ikonos images. The lower right corners of Plates 3a and 3b show a few buildings (red lines) that are completely omitted as a consequence of the classification step. Therefore, a reliable classification is a prerequisite to lead to a successful building extraction process. Third, the detection rate indicates that 64.4 percent of the

TABLE 3. QUALITY ASSESSMENT RESULTS

Steps	Branching Factor	Miss Factor	Detection Rate (Percent)	Quality Percentage
Classification	1.96	0.38	72.1	29.8
Segmentation/Squaring	0.39	0.56	64.4	51.3

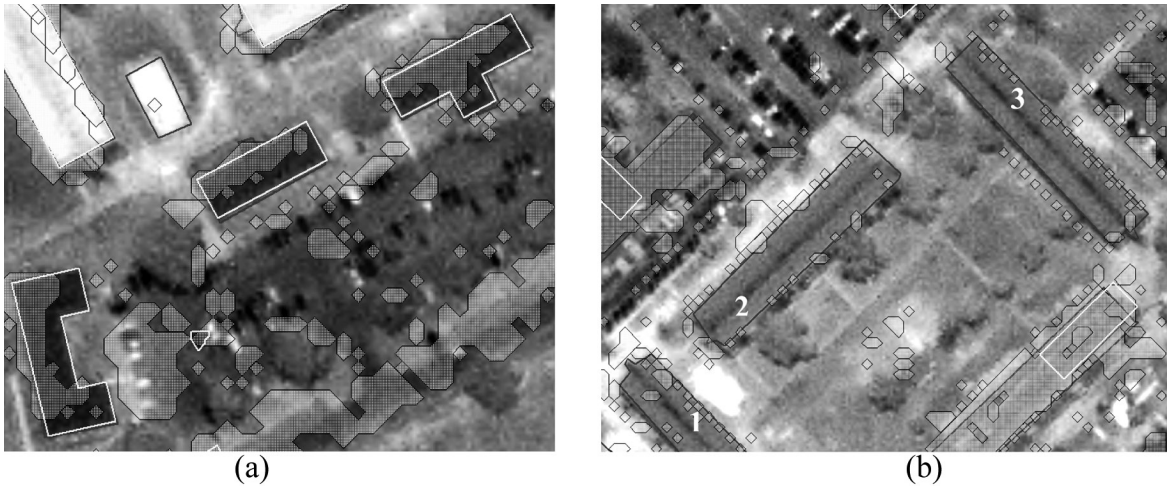


Figure 3. Successfully (a, bright line) and unsuccessfully (b) extracted buildings guided by classification (shaded area). Despite a large portion of buildings missed in the classification, the segmentation and squaring still yield satisfactory and accurate results (a). The three large buildings (labeled, dark line) are missed due to inadequate building pixels identified in the classification (b).

buildings in the image were successfully extracted by the automated approach. Missed buildings were mainly caused by the high miss factor (0.38) in the classification step as shown in Table 3. A comparison of the manual and automated results in Plate 3a also supports this observation. It should be noted that identification difficulty was encountered in the manual delineation as well because of the limit of image resolution and the lack of ground truth knowledge for the test area. As can be seen in Plate 3a, this may cause buildings, especially small ones, to be mistakenly missed or included in the manual delineation results. This will in turn boost the false positive and false negative errors in the automatic process.

In addition to the above quantitative assessment, a visual inspection was also conducted to evaluate the automatic approach. Several observations can be made based on Plates 3a and 3b. First, building edges delineated from manual and automatic approaches coincided very well. For almost all buildings, the offsets between corresponding edges were about one (1) pixel (1 meter on the ground). Second, a few larger offsets or differences were observed at building corners because of the occlusion of trees or tree shadows, which often cause unrealistically asymmetric building contours. Third, it should be noted that the low contrast and shadows caused difficulty for manual building delineation as well. A comparison with the automatic results indicates that the manual approach also yielded some unrealistic building shapes, such as non-parallel, non-perpendicular, or asymmetric building edges. This is solely limited by the resolution and quality that the Ikonos image can provide.

Conclusions

This study follows the experience gained in the past, that multiple data souces should be exploited to reach a reliable automatic solution to building extraction from digital imagery. We used building classes generated from a supervised classification to drive the automatic building extraction from Ikonos imagery. Successful results depend largely on the supervised classification to capture a comprehensive and reasonably good set of building roofs. This can be achieved by focusing the classification on the urban area, carefully selecting roof material subclasses, and picking an adequate number of training samples. Despite this effort, observable misclassification still existed between building and road classes due to the limited number of bands of the Ikonos imagery. The use of ISODATA for building segmentation avoids the difficulty in determining the segmentation threshold. The consistency check between the classification- and segmentation-derived potential building polygons can greatly reduce the false postive errors and allow for a reliable selection for candidate buidlings to be squared in the following step. The Hough transformation-based building squaring approach can effectively determine the dominant orientation of a building and extract perpendicular and parallel lines to form the building contour. This study demonstrates the efficiency of the proposed approach. Satisfactory building results with a 64.4 percent detection rate at about one pixel edge aligment accuracy were obtained in the tested urban area.

Difficulties still remain and need further efforts. The most critical problem caused by roof and road misclassification is the large number of false positives, which in turn degrades the overall performance of the entire process. Studies are needed to effectively reduce this misclassification based on semantic and geometric properties of roads and buildings. Occlusions from trees and other objects mistakenly lead to the delineation of extraneous building contours as well as omission of real building contours. The property of building symmetry needs to be introduced to trim extraneous areas and reconstruct omitted

building boundaries using shape analysis techniques. Additional work is also needed to handle circular or curved buildings, or those whose boundaries are not orthogonal.

Acknowledgment

This study was sponsored by the National Imagery and Mapping Agency (NIMA).

References

Baillard, C., and H. Maitre, 1999. 3D reconstruction of urban scenes from aerial stereo imagery: A focusing strategy, *Computer Vision and Image Understanding*, 76(3):244–258.

Baltsavias, E., M. Pateraki, and L. Zhang, 2001a. Radiometric and geometric evaluation of Ikonos GEO images and their use for 3D building modeling, *Proceedings of the Joint ISPRS Workshop "High Resolution Mapping from Space 2001,"* 19–21 September, Hannover, Germany (on CD ROM).

Baltsavias, E., A. Grün, and L.Van Gool (editors), 2001b. *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)*, A.A. Balkema Publishers, Lisse, The Netherlands, 415 p.

Biehl, L., and D. Landgrebe, 2002. MultiSpec—A tool for multispectral-hyperspectral image data analysis, *Computers & Geosciences*, 28(10):1153–1159.

Braun, C., Th. Kolbe, F. Lang, W. Schickler, V. Steinhage, A.N. Cremers, W. Förstner, and L. Plumer, 1995. Models for photogrammetric building reconstruction, *Computers & Graphics*, 19(1):109–118.

Costa, Luciano da Fontoura, and Roberto Marcondes Cesar, Jr., 2001. *Shape Analysis and Classification, Theory and Practice*, CRC Press, Boca Raton, Florida, 659 p.

Dial, G., 2000. IKONOS satellite mapping accuracy, *Proceedings of the ASPRS Annual Conference*, 22–26 May, Washington, D.C. (on CD ROM).

Dial, G., L. Gibson, and R. Poulsen, 2001. IKONOS satellite imagery and its use in automated road extraction, *Automated Extraction of Man-Made Objects from Aerial and Space Images (III)* (E. Baltsavias, A. Grün, and L. Van Gool, editors), A.A. Balkema Publishers, Lisse, The Netherlands, pp. 349–358.

Förstner, W., 1999. 3D-city models: Automatic and semiautomatic acquisition methods, *Photogrammetric Week* (D. Fritsch and R. Spiller, editors), Wichmann Verlag, Heidelberg, Germany, pp. 291–303.

Förstner, W., and L. Pluemer (editors), 1997. *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, Birkhaeuser Verlag, Berlin, Germany, 277 p.

Fraser, C., H.B. Hanley, and T. Yamakawa, 2001. Sub-metre geopositioning with Ikonos GEO imagery, *Proceedings of the Joint ISPRS Workshop "High Resolution Mapping from Space 2001,"* 19–21 September, Hannover, Germany (on CD ROM).

Fraser, C., E. Baltsavias, and A. Grün, 2002. Processing of Ikonos imagery for submetre 3D positioning and building extraction, *ISPRS Journal of Photogrammetry and Remote Sensing,* 56(3):177–194.

Fritsch, D., and R. Spiller (editors), 1999. *Photogrammetric Week*, Wichmann Verlag, Heidelberg, Germany, 374 p.

Gonzalez, R., and R. Woods, 1993. *Digital Image Processing*, Addison-Wesley Publishing Company, Reading, Massachusetts, 716 p.

Grün, A., O. Kuebler, and P. Agouris (editors), 1995. *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhaeuser Verlag, Berlin, Germany, 321 p.

Grün, A., E. Baltsavias, and O. Henricsson (editors), 1997. *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, Birkhaeuser Verlag, Berlin, Germany, 393 p.

Haala, R., 1999. Combining multiple data sources for urban data acquisition, *Photogrammetric Week* (D. Fritsch and R. Spiller, editors), Wichmann Verlag, Heidelberg, Germany, pp. 329–340.

Hofmann, P., 2001. Detecting buildings and roads from IKONOS data using additional elevation information, *GeoBIT/GIS*, 6:28–33.

Huertas, A., R. Nevatia, and D. Landgrebe, 1999. Use of hyperspectral data with intensity images for automatic building modeling, *Proceedings of the Second International Conference on Information Fusion*, 06–08 July, Sunnyvale, California (on CD ROM).

Huertas, A., Z. Kim, and R. Nevatia, 2000. Multisensor integration for building modeling, *Proceedings of Computer Vision and Pattern Recognition*, 13–15 June, Hilton Head Island, South Carolina (on CD ROM).

Kettig, R., and D. Landgrebe, 1976. Computer classification of remotely sensed multispectral image data by extraction and classification of homogeneous objects, *IEEE Transactions on Geoscience Electronics*, GE-14(1):19–26.

Kim, J.-R., and J.-P. Muller, 2001. Assessment of automated techniques for extracting vegetation and buildings from 1m stereo multispectral IKONOS and 1m pan-sharpened IKONOS coupled with scanning laser altimetry, *Proceedings of the Joint ISPRS Workshop "High Resolution Mapping from Space 2001,"* 19–21 September, Hannover, Germany (on CD ROM).

Li, R., G. Zhou, S. Yang, G. Tuell, N. Schmidt, and C. Fowler, 2000. A Study of the potential attainable geometric accuracy of IKONOS satellite imagery, *Proceedings of the ASPRS Annual Conference*, 22–26 May, Washington, D.C. (on CD ROM).

Mather, P., 1999. *Computer Processing of Remotely-Sensed Images: An Introduction, Second Edition*, John Wiley & Sons, United Kingdom, 292 p.

Mayer, H., 1999. Automatic object extraction from aerial imagery—A survey focusing on buildings, *Computer Vision and Image Understanding*, 74(2):138–149.

McKeown, D., Jr., S. Cochran, S. Ford, C. McGlone, J. Shufelt, and D.Yocum, 1999. Fusion of HYDICE hyperspectral data with panchromatic feature extraction, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1261–1277.

Muller, J.-P., J.R. Kim, and L. Tong, 2001. Automated mapping of surface roughness and landuse from simulated and spaceborne 1m data, *Automated Extraction of Man-Made Objects from Aerial Images (III)* (E. Baltsavias, A. Grün, and L. Van Gool, editors), A.A. Balkema Publishers, Lisse, The Netherlands, pp. 369–380.

Otsu, N., 1976. A threshold selection method from gray-level histograms, *IEEE Transactions of Systems, Man, and Cybernetics*, SMC-9(January):62–66.

Petrie, G., 2001. Commercial high resolution space imagery: A very long gestation and a troubled birth!!, *GeoInformatics*, 4(2):12–17.

Ray, B.K., and K.S. Ray, 1992. An algorithm for polygonal approximation of digitized curves, *Pattern Recognition Letters*, 13(7):489–496.

Sohn, G., and I.J. Dowman, 2001. Extraction of buildings from high resolution satellite data, *Automated Extraction of Man-Made Objects from Aerial Space Images (III)* (E. Baltsavias, A. Grün, and L. Van Gool, editors), A.A. Balkema Publishers, Lisse, The Netherlands, pp. 345–355.

Toutin, Th., 2001. Geometric processing of Ikonos GEO images with DEM, *Proceedings of the Joint ISPRS Workshop "High Resolution Mapping from Space 2001,"* 19–21 September, Hannover, Germany (on CD ROM).

Toutin, Th., R. Chénier, and Y. Carbonneau, 2001. 3D geometric modelling of Ikonos GEO images, *Proceedings of the Joint ISPRS Workshop "High Resolution Mapping from Space 2001,"* 19–21 September, Hannover, Germany (on CD ROM).

Weidner, U., and W. Förstner, 1995. Towards automatic building reconstruction from high resolution digital elevation models, *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(4):38–49.

(Received 06 June 2002; accepted 10 July 2002; revised 08 October 2002)